

## A System for Extensive Slash Reading Using Web

Shosaku TANAKA<sup>1</sup>, Yoichi TOMIURA<sup>2</sup>, Kensei YUKINO<sup>3</sup>

<sup>1</sup>*Computing and Communications Center, Kyushu University,*

*Hakozaki 6-10-1 Fukuoka city, Japan*

*sho@cc.kyushu-u.ac.jp*

<sup>2</sup>*Faculty of Information Science and Electrical Engineering, Kyushu University,*

*Hakozaki 6-10-1 Fukuoka city, Japan*

*tom@is.kyushu-u.ac.jp*

<sup>3</sup>*Graduated School of Information Science and Electrical Engineering, Kyushu University,*

*Hakozaki 6-10-1 Fukuoka city, Japan*

*yukino@lang.is.kyushu-u.ac.jp*

### Abstract

*This paper proposes a system for extensive slash reading using Web. Slash reading is one of the most popular and effective methods for learning languages, wherein learners read sentences with slashes at boundaries between sense groups. Since slash reading requires learners to read numerous slashed sentences, this method is a type of extensive reading. However, there are few various, level-suitable and slashed texts as materials at the moment. This research adopts texts on the Web as materials and proposes a system for the method, which can select appropriate texts and insert slashes in an arbitrary sentence. The prototype system is already working on the Internet as a Web server and can be easily accessed by anyone at any time and location.*

**Keywords:** *slash reading, extensive reading, Web, natural language processing*

## 1 Introduction

Slash reading is a popular method for learning languages. In this method, learners read sentences slashed at boundaries between sense groups as follows:

I am happy / to join with you today / in what will go down in history / as the greatest demonstration / for freedom / in the history of our nation. // Five score years ago, / a great American, / in whose symbolic shadow / we stand today, / signed the Emancipation Proclamation. // ...

(‘I have a dream’ by King, M. L.)

Learners read such a sentence per sense group, that is, ‘I am happy’, ‘to join with you’ and so on, without rearranging them into Japanese syntactical patterns. As they repeat this reading process, they get accustomed to reading and automatically comprehending an English sentence in the order of its words. It is generally considered that this method is effective not only for reading and understanding directly but also for listening. This method entails extensive reading, that is, learners are required to read numerous slashed sentences; however, materials such as text with slashes are limited. Therefore, the method has not yet been used in a classroom. In the present paper, we suggest using texts as materials, and we show the blueprint of a system for the method. Web offers several texts that could be of interest to learners and are suitable for various levels. However, there are a number of crucial issues. Most online texts are not authorized, and some may not be conducive to learning. Expectedly, most texts are not slashed, and it is difficult for learners to recognise the level of the online texts before reading it. Our proposed system comprises some modules that filter texts from those selected by learners, identify sense groups in a sentence and insert slashes. The proposed system enables learners to read a substantial amount of appropriate, interesting, level-suitable and slashed texts on the Web.

## 2 Requirements for Materials and Problems

### 2.1 Requirements for Materials for Extensive Slash Reading

As mentioned above, this method entails extensive reading, that is, learners are required to read numerous slashed sentences; therefore, this method is termed *extensive slash reading*.

(Day & Bamford, 1998) described some requirements for materials for extensive reading as follows:

- 1) A variety of materials on a wide range of topics is made available in order to encourage learners to read for different reasons and in different ways.
- 2) Learners select what they want to read and have the freedom to stop reading material that fails to interest them.
- 3) Materials are well within their linguistic competence in terms of vocabulary and grammar.
- 4) Reading can be performed individually and in silence, at their own pace and can be performed outside class at any time and location.

The current method should satisfy these requirements as well as the following:

- 5) Sentences in materials have slashes at the boundaries between sense groups.

1)-5) are the minimal requirements for materials for extensive slash reading. Although substantial material for extensive reading, such as graded readers, has been published in English, there continues to be a lack of materials and the number of slashed texts is negligible. Even if there existed plenty of materials, the cost for experts such as language teachers to slash them would be considerably high, and it would be very difficult to maintain consistency in inserting slashes because this would be based on their subjective preferences.

### 2.2 Problems of Online Texts as Materials

From the viewpoint of volume, variety and topicality, online texts appear to be preferable to authentic or published ones. Web is expanding daily, with new texts such as news or blogs, which are frequent and chronological publications of personal thoughts. However, in fact, some of these texts may not be appropriate materials because people can publish texts on Web very easily; such texts may not be grammatical and could be informal, for example private memorandums or informal messages. Learners may not recognise its appropriateness and level before reading. If learners frequently fail to read texts due to their levels of difficulty, they will lose their motivation for reading. This is one of the common problems related to extensive reading. Besides, most online texts have not been slashed. As mentioned above, Web possesses the advantage of volume, variety and topicality but has the disadvantages in terms of the quality and few texts are slashed. In the following section, we propose a system for extensive slash reading using Web in order to solve these problems by natural language processing techniques.

## 3 A System for Extensive Slash Reading Using Web

We describe the blueprint of a system for extensive slash reading using Web. Although its prototype system is already functioning on the Internet as a Web server, the system has not yet been fully completed. The current version can be accessed at the following URL:

<http://lengua.cc.kyushu-u.ac.jp/english/sr/>

### 3.1 Outline

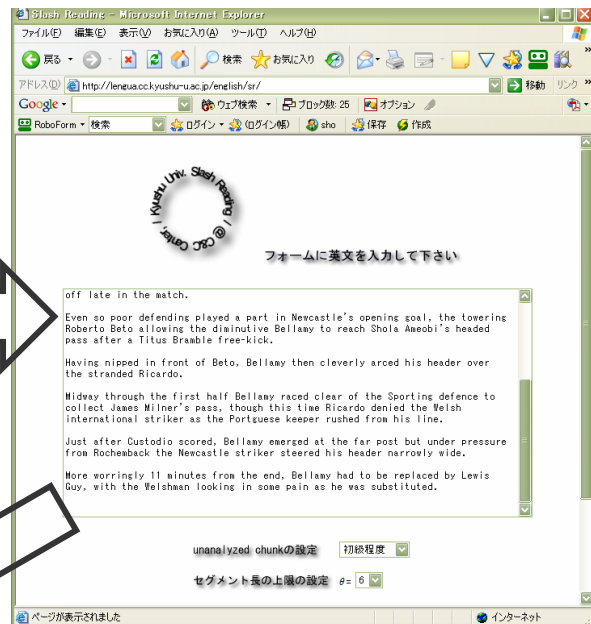
The system consists of the following three modules:

- 1) Filtering module: The first module calculates the readability of a text which learners select to ensure that it is within their linguistic competence, and its appropriateness as a material.
- 2) Slashing module: The second module identifies sense groups in the sentences and inserts slashes between them. This module uses an algorithm described in (Tanaka & Tomiura, 2004) for identifying sense groups.
- 3) Reading module: The last module is optional for learners. If learners use this module, it shows them the sentences in sense groups. Consequently, they can read the sentences in sense groups automatically, in sequential order of the words.

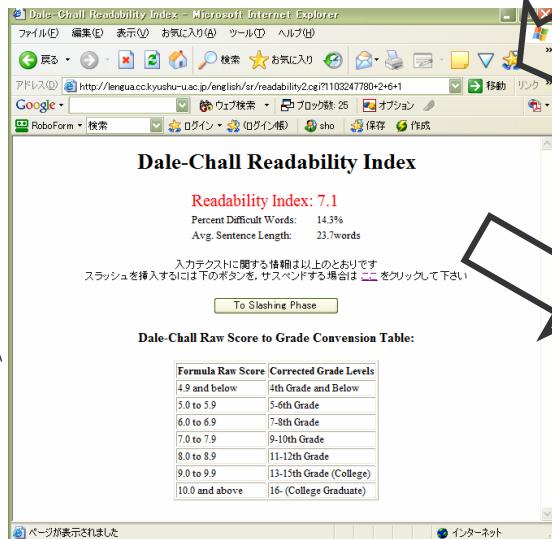
This system enables learners to perform efficient slash reading using appropriate texts on Web. Since the system is constructed as a Web server, the interface is familiar with everyone and, the learners can access it with only a Web browser at any time and location.



Learners select a text which interests them.



Learners input the text into the system.

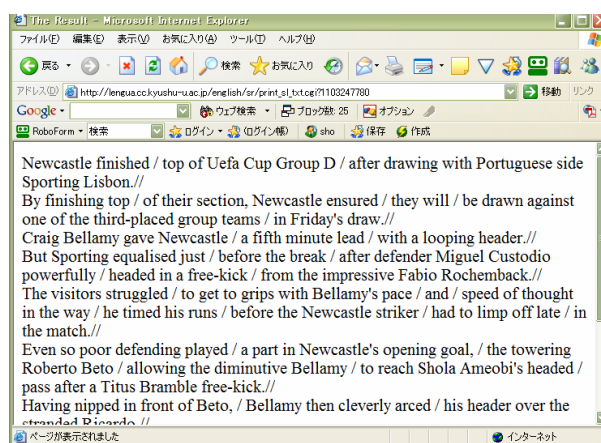


The system calculates the readability of the text, and learners determine whether they read it or not.



The system displays the sentences per sense group.

Figure 1 The prototype system.



**Figure 2** The text slashed by the system.

### 3.2 Filtering Module

The system proposed in this paper will include a module to determine the appropriateness and readability of an online text. When learners input an online text that interests them in the system, this module calculates the text's appropriateness and readability and displays the result. Learners can select not only texts of their interest but also texts that are more effective at their level of English proficiency.

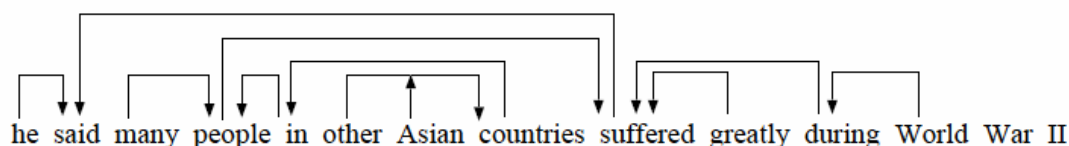
The working prototype merely calculates just only the readability of a piece of text based on the Dale-Chall readability formula (Dale & Chall, 1948). This formula can prove insufficient when applied for learners in order to determine whether they can read the text or not because it is fundamentally an index for native speakers and there is no mapping from the Dale-Chall raw score to the grade of a non-native speaker.

We suppose that the appropriateness of texts as materials is in some way related to several perspectives such as their genres, grammaticality and so on. Language teachers at Kyushu University and the authors of this paper are currently investigating such perspectives.

### 3.3 Slashing Module

The purpose of the slashing module is to identify sense groups in sentences and to insert slashes into the sentences. There are a few systematic methods for identifying sense groups and inserting slashes (Terashima, 2002). This system adopts the method proposed in (Tanaka & Tomiura, 2004)<sup>1</sup>. The method identifies sense groups in a sentence on the basis of the dependency structure of the sentence<sup>2</sup>. Suppose, for example, there is an input sentence as follows:

He said many people in other Asian countries suffered greatly during World\_War\_II (1)  
The dependency structure of sentence (1) is shown in Figure 3.



**Figure 3** The dependency structure of sentence (1).

Tanaka and Tomiura consider that a *sense group* is a subsequence in a sentence, which readers can grasp and understand for a moment. The method employs three hypotheses for identifying the sense groups:

- 1) The last word of a sense group is a headword up to its own position.
- 2) The length of a sense group is never extremely long.
- 3) The number of dependency relations across boundaries of sense groups is few.

The first hypothesis is accepted with ease. The second and the third demonstrate the tendency in sense groups. It is easy to support the second hypothesis because words in an extremely long sense group are difficult to grasp and understand directly and quickly. Therefore, the method has one parameter  $\theta$  set in advance, which is an upper limit for the length of a sense group. The third hypothesis also implies that there exist several dependency relations in a sense group. The dependency relations between words are considered to be the context of the words, which helps disambiguate their meanings. The method determines the minimum set of sense groups against the number of dependency relations across the boundaries with lengths of less than  $\theta$ .

In the case of sentence (1), words that can be the last word in a sense group are 'said', 'people', 'countries', 'suffered' and 'greatly' (excluding 'World\_War\_II' because it is the last term in the sentence).

He said many people in other Asian countries suffered greatly during World\_War\_II. (2)  
Let  $\theta$  be 6, a set of sense groups that minimise the number of dependency relations over boundaries between them is 'he said', 'many people in other Asian countries', 'suffered greatly during World\_War\_II'. As a result, we obtain the following slashed sentence<sup>2</sup>:

He said / many people in other Asian countries / suffered greatly during World\_War\_II. (3)  
The method is almost always successful in identifying the sense groups within seconds and can effectively be for applied to a large text that involves several sentences. In the prototype, although this algorithm is implemented by Perl<sup>3</sup>, most sentences can be slashed in a few seconds.

### 3.4 Reading module

This is an optional module, which is used when learners read a slashed sentence. In slash reading, learners are required to discipline themselves in order to read linearly in the manner of native speakers. This is cumbersome for learners, particularly for beginners. Therefore, this module regulates their reading process by only displaying a sentence sense group by sense group. Learners can read a sentence per sense group without deliberate effort by using this module.

For example, the sentence (3) is shown to the learners as follows:

He said  
↓  
many people in other Asian countries  
↓  
suffered greatly during World\_War\_II

In the prototype, the interval time of for which one sense group is displayed is constant, namely it is independent of the length of a sense group. The timeframe should definitely be varied according to the length of a sense group<sup>6</sup>. Moreover, it is considered that the physical length, which refers to the number of letters in a sense group, should also be an important factor determining the timeframe.

## 4 Conclusion

We have demonstrated the problems of using texts on the Web as materials for extensive slash reading, and present an outline of a system for effectively carrying out this process. As mentioned above, the filtering module and the reading module have not been fully completed. In order to complete the former module, we have to investigate the factors related to effective learning, which is one of the most interesting aspects of extensive reading. In order to complete the latter module, it will be useful to refer to the study on reading process in cognition science. Furthermore, the evaluation of this system is the most important future task. A project with an English language teacher at the Kyushu University (P&P)

has been launched this year<sup>4</sup>. We plan to attempt to use our system as a part of this and evaluate its effect on learners.

### Acknowledgements

We would like to thank Shohei Yamamoto, Ryota Yakushiji and the referees for providing us with valuable advice.

### Notes

1. (Tanaka & Tomiura, 2004) tried the preliminary experiment based on non-native speakers' introspection of sense groups, and got its accuracy of 79.4%. Recently, a new algorithm has been proposed, which considers both dependency relations and syntactic categories between the relations, such as S-complement, TO-infinitive and so on.
2. Sense groups are based on not only syntactic information but also semantic information. However, it is difficult to deal with such semantic information in the deterministic and systematic way automatically for the present.
3. In the set, the number of dependency relations over boundaries between sense groups is two ('suffered→said' and 'people→suffered').
4. Perl is one of the general-purpose programming languages. Because Perl programs are executed by the Perl interpreter, the speed of execution is not so fast.
5. This project involves the establishment of infrastructure for Computer Assisted Language Learning (CALL).
6. From the viewpoint of extensive reading in (Day & Bamford, 1998), it is important for learners to read a text at their own pace. Therefore, we should consider displaying sense groups by themselves on the system.

### References

- Dale, E. and Chall, J. 1948. *A formula for predicting readability*. Educational Research Bulletin, 27, 11-20, pp.37-54.
- Day, R. R. and Bamford, J. 1998. *Extensive Reading in the Second Language Classroom*. Cambridge University Press.
- Tanaka, S. and Tomiura, Y. 2004. *Construction of the System for Slash Reading -An Algorithm for Slashing Based on the Chunking Process-*. Proceeding of Workshop 'Natural Language Processing for e-Learning', Annual Conference of Natural Language Processing, pp. 37-40. (In Japanese)
- Terashima, M. 2002. *A New Way of Phrase Reading*. Asunaro Press. (In Japanese)